

Stochastic Optimization Models for a Single-Sink Transportation Problem

Francesca Maggioni*

Michal Kaut[†]

Luca Bertazzi[‡]

Abstract

In this paper we study a single-sink transportation problem in which the production capacity of the suppliers and the demand of the single customer are stochastic. Shipments are performed by capacitated vehicles, which have to be booked in advance, before the realization of the production capacity and the demand. Once the production capacity and the demand are revealed, there is an option to cancel some of the booked vehicles against a cancellation fee. If the quantity shipped from the suppliers using the booked vehicles is not enough to satisfy the demand, the residual quantity is purchased from an external company. The problem is to determine the number of vehicles to book in order to minimize the total cost.

We formulate a two-stage and a multistage stochastic mixed integer linear programming models to solve this problem and test them on a real case provided by *Italcementi*, the primary Italian clinker producer and the 5th largest cement producer in the world. We test the influence of different scenario-tree structures on the solutions of the problem, as well as sensitivity of the results with respect to the cancellation fee.

Key Words. Single-sink transportation problem, stochastic programming, scenario tree.

Introduction

In the classical *Transportation Problem* a set of retailers is served by a set of suppliers. Each retailer has a deterministic demand, while each producer has a given production capacity. A transportation cost is charged for each unit of product sent from the suppliers to the retailers. The problem is to determine the quantity of product to send from each supplier to each retailer in order to minimize the total transportation cost. A particular case is given by the so called *Single-Sink Transportation Problem*, in which a single retailer is served by a set of suppliers. This problem has been deeply studied, in particular when the total cost is given by the sum of a variable transportation cost and of a fixed charge cost to use the supplier – see Alidaee and Kochenberger (2005), Herer, Rosenblatt, and Hefter (1996), Lamar and Wallace (1997), Lamar, Sheffi, and Powell (1990) and Palekar, Karwan, and Zionts (1990). The reason is that this problem, referred to as *Single-Sink Fixed-Charge Transportation Problem*, has several applications in Logistics, including for instance the supplier selection problem and the fleet selection problem.

The transportation problem we study has been inspired by a real case of *clinker* replenishment, provided by the largest Italian cement producer located in Sicily, the large island adjacent to the

*Department of Mathematics, Statistic, Computer Science and Applications, Bergamo University, Via dei Caniana 2, Bergamo 24127, Italy, e-mail: francesca.maggioni@unibg.it

[†]Molde University College, P.O. Box 2110, N-6402 Molde, e-mail: michal-kaut@hiMolde.no

[‡]Department of Quantitative Methods, Brescia University, Contrada S. Chiara 50, Brescia 25122, Italy, e-mail: bertazzi@eco.unibs.it

southern tip of Italy. Clinker is a semi-worked material which is ground for use in cement and then sold to final customers. The logistic system we study is organized as follows (see Figure 1): The *clinker* is produced by production plants located both in Sicily and in Calabria, the neighbouring region of mainland Italy. The four production plants considered in the model are Isola delle Femmine (Palermo, Sicily), Porto Empedocle (Agrigento, Sicily), Castrovillari (Cosenza, Calabria) and Vibo Valentia (Calabria). The warehouse is located in Catania, Sicily. Since the transportation from Calabria requires a ferryboat, in addition to trucks, the corresponding transportation costs are higher than those from inside Sicily. The production capacity of the four production plants, as well as the demand for clinker at Catania, is considered stochastic.



Figure 1: Sicily and Calabria.

All the vehicles are leased from an external transportation company, which we assume to have an unlimited fleet. The vehicles must be booked in advance, before the demand and production capacity are revealed. Only full load shipments are allowed. When the demand and the production capacity become known, there is an option to cancel some of the reservations against a cancellation fee. If the quantity delivered from the four suppliers using the booked vehicles is not enough to satisfy the demand in Catania, the residual quantity is purchased from an external company at a higher price. The problem is to determine, for each supplier, the number of vehicles to book in order to minimize the total cost, given by the sum of the transportation cost (including the cancellation fee for vehicles booked but not used) and the cost of the product purchased from the external company.

This problem can be seen as a stochastic single-sink transportation problem. The stochastic pa-

rameters are the demand and the production capacity of the production plants. A simple solution to this problem could be obtained by solving the so called *Expected value problem* or *Mean value problem*, that is the corresponding deterministic problem in which the stochastic parameters are replaced by their mean value. We propose two stochastic mixed integer linear programming models for the solution to this problem. We start with a two-stage stochastic programming model with recourse, see for instance Birge and Louveaux (1997), Kall and Wallace (1994); the drawback of this model is represented by the so-called *end-of-horizon effects*, characterized by the fact that warehouse is typically empty at the end of the second stage. To mitigate the problem, we propose a multistage stochastic model—see again Birge and Louveaux (1997) or Kall and Wallace (1994) for more details.

These two models allow us to solve the real case and to carry out computational experiments in order to answer the following questions: How bad is the solution obtained by solving the corresponding deterministic problem? How many extra periods do we need to mitigate the end-of-horizon effects? Does the model benefit from having a scenario tree with more than one stage with branching? How does the optimal solution vary by modifying the cancellation fee?

The rest of the paper is organized as follow. In Section 1 we formally describe the problem. In Section 2 we formulate the corresponding two-stage and multistage stochastic mixed integer linear programming models. In Section 3 we present computational results and answer the above questions.

1 Problem description

A set $\mathcal{I} = \{i : i = 1, \dots, I\}$ of suppliers has to satisfy the demand of a single customer. The initial inventory level at the customer is l_0 , the storage capacity at the customer is l_{\max} and the unloading capacity at the customer is g . Each supplier i is characterised by its production cost c_i and transportation cost t_i . The price for buying the clinker from an external source is denoted b and we assume $b > \max_i(c_i + t_i)$. A set $\mathcal{S} = \{s : s = 1, \dots, S\}$ of possible scenarios is available. Each scenario $s \in \mathcal{S}$ is described by its probability p^s , the demand d^s of the customer and the production capacity a_i^s of each supplier $i \in \mathcal{I}$. We assume a uniform fleet of vehicles with capacity q and allow only full-load shipments. The cancellation fee is given as a proportion α , $0 \leq \alpha \leq 1$, of the transportation costs t_i , so the transportation cost of each vehicle from the supplier i is qt_i if the vehicle is booked and then used, or αqt_i if the vehicle is booked, but later cancelled.

The problem is to determine the number of vehicles $X_i \in \mathbb{N}$ to book for transport from each supplier $i \in \mathcal{I}$ before the demand and the production capacity are revealed (first stage decision variables), the number of vehicles $Z_i^s \in \mathbb{N}$ actually used in scenario $s \in \mathcal{S}$ (second stage variables) and, finally, the quantity Y^s , $Y^s \geq 0$, of product to purchase from an external source in scenario $s \in \mathcal{S}$ (second stage variable). The aim is to minimize the total cost, given by the sum of the transportation cost (including the cancellation fee for vehicles booked, but not used) and the cost of the product purchased from the external source.

2 Model formulation

2.1 The two-stage stochastic model

In the two-stage (one-period) case, we get the classical stochastic programming model with recourse:

$$\min q \sum_{i=1}^I t_i X_i + \sum_{s=1}^S p^s \left[b Y^s - (1 - \alpha) q \sum_{i=1}^I t_i (X_i - Z_i^s) \right] \quad (1)$$

subject to

$$q \sum_{i=1}^I X_i \leq g \quad (2)$$

$$l_0 + \sum_{i=1}^I qZ_i^s + Y^s - d^s \geq 0, \quad \forall s \in \mathcal{S} \quad (3)$$

$$l_0 + \sum_{i=1}^I qZ_i^s + Y^s - d^s \leq l_{\max}, \quad \forall s \in \mathcal{S} \quad (4)$$

$$Z_i^s \leq X_i, \quad \forall i \in \mathcal{I}, \forall s \in \mathcal{S} \quad (5)$$

$$qZ_i^s \leq a_i^s, \quad \forall i \in \mathcal{I}, \forall s \in \mathcal{S} \quad (6)$$

$$X_i \in \mathbb{N}, \quad \forall i \in \mathcal{I} \quad (7)$$

$$Y^s \geq 0, \quad \forall s \in \mathcal{S} \quad (8)$$

$$Z_i^s \in \mathbb{N}, \quad \forall i \in \mathcal{I}, \forall s \in \mathcal{S}. \quad (9)$$

The first sum in the objective function (1) is the booking cost of the vehicles, while the second sum represents the recourse action, consisting of buying extra clinker (Y^s) and cancelling unwanted vehicles. Note that cancelling a vehicle results in a negative cost, since we get back the price paid in stage one, minus the cancellation fee. As a result, it is possible to have negative second-stage costs. Constraint (2) guarantees that the total quantity delivered from the suppliers to the customer is not greater than the customer's unloading capacity g , inducing thus an upper bound on the total number of vehicles. Constraints (3) and (4) ensure that the second-stage storage level is between zero and l_{\max} . Constraint (5) guarantees that the number of vehicles servicing supplier i is at most equal to the number booked advanced and (6) controls that the quantity of clinker delivered from supplier i does not exceed its production capacity a_i^s . Finally, (7)–(9) define the first- and second-stage decision variables of the problem.

The drawback of the two-stage stochastic model is that can be expected to exhibit a strong end-of-horizon effect in the form of an empty warehouse at the end of the second stage. This is due to the fact that we do not assign any value to clinker in the storage and, at the same time, it is possible to cancel vehicles that are not needed to satisfy the demand.

2.2 The multistage stochastic model

We now formulate a multistage stochastic model. The aim of the model is the same, that is to determine, for each supplier, the number of vehicles to book in order to minimize the total cost.

The major difference between two-stage and multistage models is that in the latter, we have to model the structure of the underlying scenario tree, see for example Dupačová (1997), Dupačová, Consigli, and Wallace (2000). This can be done in two ways: we can either model the problem by scenarios and then add a set of so-called *non-anticipativity constraints*, or write the model in terms of *nodes* of the scenario tree and describe the tree structure by giving to each node (except the root of the tree) a pointer to its parent, i.e. the node immediately preceding it. We use the latter formulation and need thus the following notation, in addition to the one introduced in Section 1:

- $\mathcal{N} = \{n : n = 0, \dots, N\}$ is the ordered set of nodes of the scenario tree structure;
- $0 \in \mathcal{N}$ is the root of the tree, which represents the time instant for which we want to determine the number of vehicles to book;

- $\mathcal{F} = \{n : n = N - F + 1, \dots, N\} \subset \mathcal{N}$ is the set of the leaves of the tree, that is the nodes in the last stage of the model; since the *number of scenarios* S is equal to the number of leaves, we get $S = F = |\mathcal{F}|$;
- $\text{pa}(n)$ denotes the parent of the node $n \in \mathcal{N} \setminus \{0\}$ in the scenario tree;
- p^n is the probability of node $n \in \mathcal{N}$; we have considered the leaves (and therefore the scenarios) to be equiprobable, so that

$$p^n = \begin{cases} \frac{1}{|\mathcal{F}|} & \text{if } n \in \mathcal{F} \\ \sum_{\substack{m \in \mathcal{N} \setminus \{0\} \\ \text{pa}(m)=n}} p^m & \text{if } n \in \mathcal{N} \setminus \mathcal{F}; \end{cases} \quad (10)$$

- Parameters a_i^n and d^n correspond to a_i^s and d^s from the previous section; they are both defined at nodes $n \in \mathcal{N} \setminus \{0\}$;
- Variable $X_i^n \in \mathbb{N}$, defined at $n \in \mathcal{N} \setminus \mathcal{F}$, corresponds to X_i^s from the previous section;
- Variables Y^n and Z_i^n correspond to Y^s and Z_i^s and are thus defined on $n \in \mathcal{N} \setminus \{0\}$;
- L^n is a new variable representing the inventory level of the customer at node n :

$$L^n = L^{\text{pa}(n)} + q \sum_{i=1}^I Z_i^n + Y^n - d^n, \quad \forall n \in \mathcal{N} \setminus \{0\}; \quad (11)$$

The multistage model can be then formulated as follows:

$$\min \sum_{n=0}^{N-F} p^n \left[q \sum_{i=1}^I t_i X_i^n \right] + \sum_{n=1}^N p^n \left[b Y^n - (1 - \alpha) q \sum_{i=1}^I t_i \left(X_i^{\text{pa}(n)} - Z_i^n \right) \right] \quad (12)$$

subject to

$$q \sum_{i=1}^I X_i^n \leq g, \quad \forall n \in \mathcal{N} \setminus \mathcal{F} \quad (13)$$

$$L^{\text{pa}(n)} + q \sum_{i=1}^I Z_i^n + Y^n - d^n \geq 0, \quad \forall n \in \mathcal{N} \setminus \{0\} \quad (14)$$

$$L^{\text{pa}(n)} + q \sum_{i=1}^I Z_i^n + Y^n - d^n \leq l_{\max}, \quad \forall n \in \mathcal{N} \setminus \{0\} \quad (15)$$

$$Z_i^n \leq X_i^{\text{pa}(n)}, \quad \forall i \in \mathcal{I}, \forall n \in \mathcal{N} \setminus \{0\} \quad (16)$$

$$q Z_i^n \leq a_i^n, \quad \forall i \in \mathcal{I}, \forall n \in \mathcal{N} \setminus \{0\} \quad (17)$$

$$X_i^n \in \mathbb{N}, \quad \forall i \in \mathcal{I}, \forall n \in \mathcal{N} \setminus \mathcal{F} \quad (18)$$

$$Y^n \geq 0, \quad \forall n \in \mathcal{N} \setminus \{0\} \quad (19)$$

$$Z_i^n \in \mathbb{N}, \quad \forall i \in \mathcal{I}, \forall n \in \mathcal{N} \setminus \{0\}, \quad (20)$$

where (12)–(20) correspond one-to-one to (1)–(9) in Section 2.1. Even if there is no time explicitly in the model, we will need to refer to it later in the paper. For this purpose, we denote by T the number of stages in the model and introduce a time index $t = 1 \dots, T$, with $t = 1$ corresponding to the root of the scenario tree.

Note that the two-stage model presented in the previous section is a special case of the above multistage model. The reason we present both formulations is that the two-stage model may be easier to understand, especially for readers not used to the node-based formulation of a stochastic program.

3 Numerical results

In this section, we present numerical results from the case described in Introduction, based on historical data from period January 2003 to May 2007 presented in Tables 1–3: Table 1 presents the production and transportation costs for each supplier, together with its distance from the customer in Catania; Table 2 shows the historical monthly data of demand for clinker in the area of Catania and, finally, Table 3 reports the monthly production capacity of each supplier in the considered period.

Table 1: Production costs c_i , transportation costs t_i , distance and travel time from Catania.

Supplier	c_i (€/t)	t_i (€/t)	Dist. (km)	Time (h)
Porto Empedocle (AG)	18.79	11.40	180	3
Castrovillari (CS)	9.55	33.00	356	7
Isola d. Femmine (PA)	11.00	14.10	225	3
Vibo Valentia (VV)	11.54	18.50	181	4

Table 2: Monthly demand d of Catania in the period from January 2003 to May 2007, measured in kilotonnes.

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2003	20.0	23.3	23.7	24.7	24.4	20.4	24.3	22.1	19.6	26.9	26.4	22.1
2004	24.7	24.5	30.8	24.7	29.9	21.2	23.8	21.0	18.1	28.8	23.1	24.0
2005	24.4	23.8	27.7	28.6	37.4	29.9	29.7	15.4	20.1	25.4	22.9	23.6
2006	22.8	24.6	28.9	23.6	30.9	25.4	29.6	23.3	25.3	34.4	25.7	25.4
2007	25.0	29.0	26.0	23.0	28.0							

In our computational experiment, we have used vehicle capacity $q = 30$ tonnes (t), the storage capacity $l_{max} = 35$ kilotonnes (kt) and the *daily* unloading capacity of 1800 t, giving us the *monthly* unloading capacity $g = 21 \times 1800 \text{ t} = 37.8 \text{ kt}$, or 1260 full vehicles. The cost of clinker from an external source was set to $b = \text{€}45/\text{t}$ and the cancellation fee to $\alpha = 0.5$. For the initial inventory level l_0 at the customer, we have taken the value at the beginning of January 2007, that is $l_0 = 2000 \text{ t}$, even if we are aware that it is quite small.

Our aim is to find, for each supplier, the number of vehicles to book at the beginning of January. Because of the limited number of data at our disposal, we cannot describe the stochastic parameters by a particular probability distribution; instead, we have constructed scenarios by using the historical data directly, using all the values for December, January and February as future scenarios for the January demand and production capacity. In this way, we get fourteen scenarios in the second stage of the scenario tree.

Table 3: Monthly production capacity a_i of the suppliers $i \in \mathcal{I}$ in the period from January 2003 to May 2007, measured in kilotonnes. The zero entries represent production site closures due to equipment failure and/or maintenance.

i	Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
AG	'03	9.1	4.0	11.1	14.6	21.7	14.2	17.4	8.4	24.9	17.4	12.3	13.0
	'04	0.0	4.1	9.0	10.5	9.3	12.2	11.6	13.6	9.4	11.0	9.7	0.0
	'05	0.0	9.1	8.3	21.1	15.0	15.1	12.1	13.2	11.3	13.0	7.1	1.2
	'06	1.7	9.5	4.5	14.0	12.5	15.2	11.3	15.9	6.2	11.9	7.2	9.0
	'07	13.0	13.0	19.0	4.0	10.0							
CS	'03	10.9	14.0	13.9	19.1	14.1	13.0	4.5	0.0	4.0	13.7	9.1	4.5
	'04	8.3	6.3	3.0	0.0	16.2	14.2	12.3	14.4	19.8	19.3	20.0	15.2
	'05	15.1	10.8	21.9	19.7	15.3	10.8	6.3	0.0	9.1	23.2	11.7	0.9
	'06	18.7	0.0	8.9	16.0	17.6	13.9	4.8	5.0	14.1	24.3	14.5	8.1
	'07	17.0	8.0	0.0	0.0	10.0							
PA	'03	15.5	18.1	23.3	12.4	0.5	5.7	12.5	13.5	12.3	10.2	8.3	12.0
	'04	27.1	10.0	12.8	13.8	13.7	14.0	10.6	1.4	10.3	12.6	11.5	16.9
	'05	16.0	3.8	10.6	16.6	23.0	27.7	16.7	13.4	16.8	11.1	19.0	22.4
	'06	27.5	21.5	18.6	20.4	0.0	14.0	14.3	11.2	18.4	16.9	9.4	11.1
	'07	11.0	9.0	7.0	6.0	10.0							
VV	'03	4.9	1.2	12.7	2.7	19.3	11.9	5.4	3.0	14.6	3.4	15.2	2.5
	'04	4.0	9.4	18.3	10.5	13.9	8.6	6.2	4.3	7.2	12.4	9.5	0.0
	'05	3.5	21.1	20.8	13.0	23.5	19.1	8.2	8.6	4.6	9.2	16.2	16.0
	'06	8.5	22.3	21.7	15.1	7.4	10.3	0.0	2.5	4.3	5.2	18.3	6.3
	'07	0.0	0.0	0.0	0.0	10.0							

The problem was modelled in AMPL and solved using the CPLEX solver. Since we had to use very small scenario trees, the solution time is negligible even if the problem is a stochastic mixed integer program.

In the rest of this section, we present several tests, the goal of which can be summarized as follows:

- Test the importance of the number of *proper stages*, that is stages with branching, on the quality of the first stage solution;
- Test, for each type of tree, how many *extra-periods*, that is stages without branching, we need to add to mitigate the end-of-horizon effects.

3.1 Effect of extra periods on a model with one proper stage

First, we test the model with scenario trees \mathcal{T}_1 – \mathcal{T}_3 , presented in Figures 2–4. All the trees are characterized by a branching of fourteen descendants from the root and a unique descendant on all the other nodes, such that $S = |\mathcal{F}| = 14$. The aim of the test is to determine how many *extra-periods* we need to add to the scenario tree \mathcal{T}_1 to overcome the end-of-horizon effects.

The results are reported in Table 4, where the last two columns show the cost incurred in the root

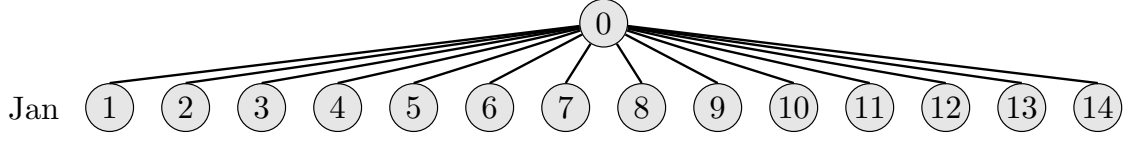


Figure 2: Scenario tree \mathcal{T}_1 .

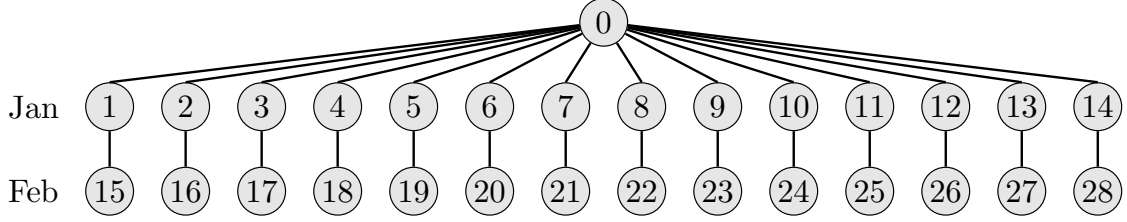


Figure 3: Scenario tree \mathcal{T}_2 .

and in the first two stages, that is

$$\text{root: } q \sum_{i=1}^4 t_i X_i^0 \quad (21)$$

$$\text{stages 1+2: } q \sum_{i=1}^4 t_i X_i^0 + \sum_{n=1}^{14} p^n \left[b Y^n - (1 - \alpha) q \sum_{i=1}^4 t_i (X_i^0 - Z_i^n) \right]. \quad (22)$$

Table 4: Solution to the tests from Section 3.1. For each scenario tree considered, the table shows the number of scenarios S and stages T , the optimal number of booked vehicles X_i for each supplier $i \in \mathcal{I}$, and the optimal costs in the root and in the first two stages, respectively.

tree	S	T	AG	CS	PA	VV	cost 1 (€)	cost 1+2 (€)
\mathcal{T}_1	14	2	400	0	563	117	439 884	438 304
\mathcal{T}_2	14	3	433	0	563	133	460 050	459 350
\mathcal{T}_3	14	7	433	0	563	133	460 050	459 446

We see that using the simplest scenario tree \mathcal{T}_1 leads to significantly fewer booked vehicles, demonstrating nicely the end-of-horizon problem. Since there is no difference between the first-stage solutions from trees \mathcal{T}_2 and \mathcal{T}_3 , we can conclude that adding just one extra period is enough to mitigate the end-of-horizon effect and achieve stability of the first-stage solution. Note, however, that this is not true for the second-stage solution, which is slightly cheaper in tree \mathcal{T}_2 than in \mathcal{T}_3 ; the reason is that in \mathcal{T}_2 the second period is the last one, so we again get the end-of-horizon effect of empty warehouse at the end of the period.

Note also that, for all the three trees, the cost of the first and second stage is actually smaller than the cost of the first stage only. This is due to the fact that the costs of buying extra clinker is smaller than the repayment we get for cancelling orders for some of the vehicles – see (22).

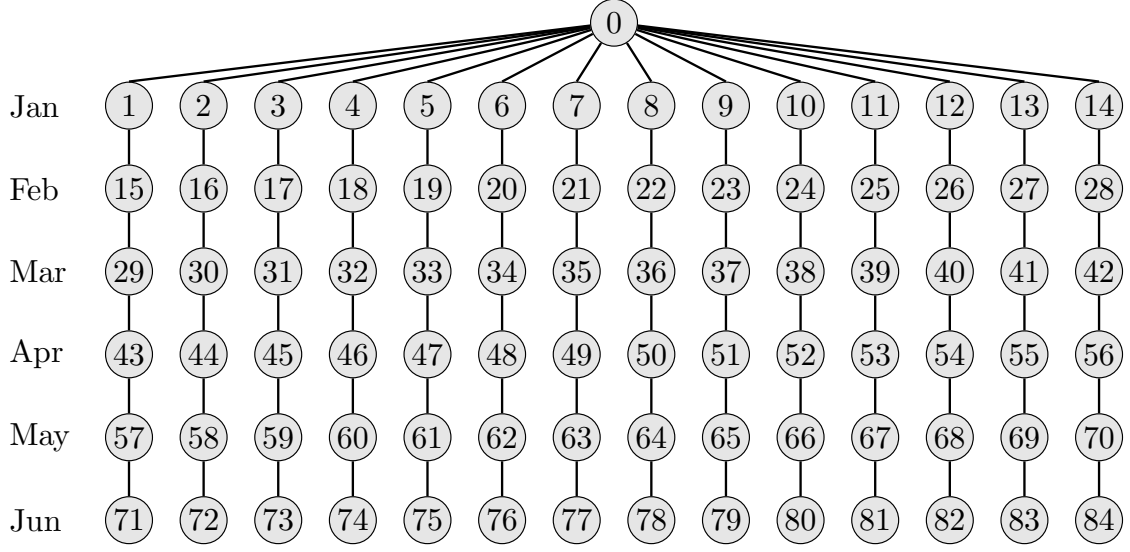


Figure 4: Scenario tree T_3 .

3.2 Deterministic solutions and the value of stochastic solution

To check the importance of modelling the randomness of the parameters, we compare the optimal solutions and objective values of the stochastic models with those obtained from the corresponding *deterministic* models, where the stochastic parameters (represented by the demand d^n and the production capacity a_i^n , $i \in \mathcal{I}$, $n \in \mathcal{N} \setminus \{0\}$) are replaced by their mean values, computed over all the nodes at the same time period. In literature, this kind of problem is called *Expected value problem* or *Mean value problem*, (see Birge and Louveaux (1997) and Kall and Wallace (1994)).

In a deterministic problem the future is completely known, so the model will always book the exact numbers of vehicles needed for the next period. In other words, we get $Z_i^n = X_i^{\text{pa}(n)}$ for all $n \in \mathcal{N} \setminus \mathcal{F}$. Furthermore, as long as we have enough transportation capacity, the model will never purchase extra clinker, i.e. $Y^n = 0$ at all $n \in \mathcal{N} \setminus \mathcal{F}$.

Solutions to the deterministic models with two, three and four stages are reported in Table 5. As expected, the model simply sorts the suppliers according to the transportation costs and books a full production capacity from the cheapest one (AG), following by the next-cheapest (PA). Just as in the stochastic version, the one-period model orders less than the multi-period versions, though the difference is much smaller than in the stochastic case.

Table 5: Solution to the tests from Section 3.2. Optimal solutions of the deterministic models with varying number of stages T . For each T , the table shows the optimal number of booked vehicles X_i for each supplier $i \in \mathcal{I}$ and the optimal costs in the root and in the first two stages, respectively.

T	AG	CS	PA	VV	cost 1 (€)	costs 1+2 (€)
2	206	0	530	0	294 642	294 898
3	206	0	533	0	295 911	295 911
4	206	0	533	0	295 911	295 911

Let's now focus on the two-stage case. We see that the deterministic model books much fewer

vehicles than the stochastic one, resulting in a first-stage solution that is only two-thirds the cost of the stochastic counterpart. However, we have to remember that this is an *in-sample* objective value (using the terminology from Kaut and Wallace (2007)) and the true cost of the solution—or the *out-of-sample* objective value—is likely to be higher. To see how much, we can solve the stochastic model with tree \mathcal{T}_1 and the first-stage variables (X_i^0) fixed to the deterministic solution. This results in a total cost of €495 788, much higher than the predicted (in-sample) cost of €294 898. Furthermore, we see that the resulting total cost is significantly higher than the optimal solution for tree \mathcal{T}_1 . The difference is known as the *Value of stochastic solution* (VSS), see e.g. Birge and Louveaux (1997). In our case, it is

$$\begin{aligned} \text{VSS} &= \text{obj. val.}(\text{deterministic sol. on } \mathcal{T}_1) - \text{obj. val.}(\text{optimal sol. of } \mathcal{T}_1) \\ &= 495\,788 - 438\,304 = 57\,384. \end{aligned} \quad (23)$$

This shows that one can save about 12% of the cost by using even the simplest stochastic model, compared to the deterministic model—all this provided that \mathcal{T}_1 is a true representation of reality. While this is certainly not the case, \mathcal{T}_1 is still a better approximation than the deterministic model. In the next section, we improve the approximation further by adding one more proper stage.

3.3 One versus two proper stages

In this section, we test the effect of one extra *proper stage* (i.e. stage with branching) on the quality of the first-stage solution. For this purpose, we consider the scenario tree \mathcal{T}_4 from Figure 5. This is a three stage tree with 14 branches from the root and 5 from each of the second-stage nodes, resulting in $S = 14 \times 5 = 70$ scenarios and $|\mathcal{N}| = 85$ nodes. The first two stages are equal to trees \mathcal{T}_1 – \mathcal{T}_3 .

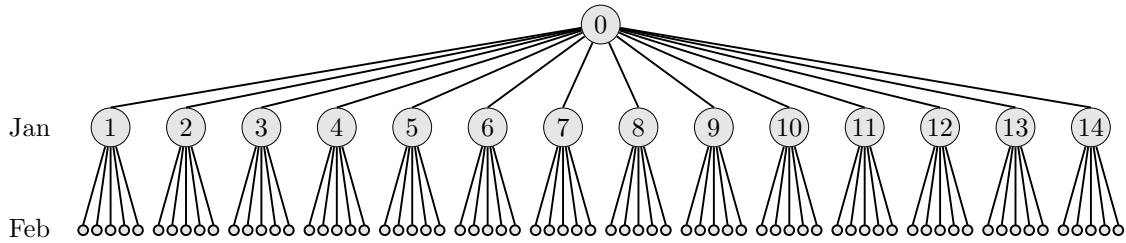


Figure 5: Scenario tree \mathcal{T}_4 .

To compare this tree with \mathcal{T}_2 (which has the same number of periods), we need to ensure that the smaller tree is an unbiased approximation of \mathcal{T}_4 . To do this, we set the last-stage values of a_i^n and d^n in the smaller tree equal to the expected value of the five corresponding last-stage values from \mathcal{T}_4 . We denote the resulting tree by $\mathcal{T}_2^{\text{ev}}$.

The results for the two scenario trees are presented in Table 6. Analogously to the previous tests, we see that a better description of stochasticity leads to larger bookings in the first stage. Actually, in the case of tree \mathcal{T}_4 the total number of booked vehicles is equal to 1260, that is the customer's unloading capacity. This is due to the fact that we use a rather low initial inventory level $I_0 = 2000$, this being the actual case from January 2007.

Just as in the previous section, we have to remember that the values presented in Table 6 represent *in-sample* values, so the costs are not directly comparable. To be able to estimate the effect of using a better scenario tree, we have to compare the *out-of-sample* costs. For this purpose, we declare \mathcal{T}_4 to be the true representation of the real world and use it as a *benchmark* to evaluate the cost of optimal

Table 6: Solution to the in-sample test in Section 3.3. For both scenario trees, the table shows the number of scenarios S and stages T , the optimal number of booked vehicles X_i for each supplier $i \in \mathcal{I}$, and the optimal costs in the root and in the first two stages, respectively.

tree	S	T	AG	CS	PA	VV	cost 1 (€)	costs 1+2 (€)
$\mathcal{T}_2^{\text{ev}}$	14	3	400	0	563	133	448 764	455 744
\mathcal{T}_4	70	3	426	21	603	210	538 101	494 894

solutions obtained using other trees. The solutions are evaluated on a *rolling-horizon* basis, that is we start by rooting each of the tested trees at node 0 of tree \mathcal{T}_4 , solve the model, fix the first-stage solution and then repeat the procedure with trees rooted at the second-stage nodes 1–14 of tree \mathcal{T}_4 .

Now, since the model does not have any constraints that would “cut across” the scenarios, solving the model on tree \mathcal{T}_4 with all the root variables fixed is equivalent to solving fourteen two-stage models, conditional on the fixed values at node 0. This simplifies the testing, as we can solve all the trees rooted at nodes 1–14 at once, by solving the model on tree \mathcal{T}_4 . The details of the procedure are provided separately for every scenario-tree structure tested:

$\mathcal{T}_2^{\text{ev}}$ tree This can be tested exactly in the way described above, that is

1. Solve the model on the $\mathcal{T}_2^{\text{ev}}$ tree and store values of the first-stage (root) variables X_i^0 .
2. Solve the model on tree \mathcal{T}_4 , with X_i^0 fixed to the stored values.

Note however, that this test actually simulates a situation where we use only a one-period model at $t = 2$, so it can be expected to underestimate the true quality of the tree. This is, however, an unavoidable consequence of having \mathcal{T}_4 as a benchmark tree.

\mathcal{T}_1 tree This case is analogous to the previous one, this time without any problems at $t = 2$.

1. Solve the model on the \mathcal{T}_1 tree and store values of the first-stage (root) variables X_i^0 .
2. Solve the model on tree \mathcal{T}_4 , with X_i^0 fixed to the stored values.

deterministic This case differs from the previous one in that we can not use tree \mathcal{T}_4 in Step 2 above, as this would simulate a situation where we use a deterministic model in at $t = 1$, but a stochastic (one-period) model at $t = 2$ and therefore overestimate the quality of the deterministic approach. Instead, we have to use the $\mathcal{T}_2^{\text{ev}}$ tree at $t = 2$, since it has a deterministic second period:

1. Solve the deterministic model and store the solution X_i^0 .
2. Solve the model on tree $\mathcal{T}_2^{\text{ev}}$, with X_i^0 fixed to the stored values. Store all the second-stage variables.
3. Finally, solve the model on tree \mathcal{T}_4 , with all the first- and second-stage variables fixed to the stored values.

Results of the tests are presented in Table 7. We see that if \mathcal{T}_4 was the correct representation of the world, there would be significant advantage to using a even the simplest one-period stochastic-programming approach, instead of solving the deterministic problem on the rolling-horizon basis. Furthermore, having two proper stages (\mathcal{T}_4) is better than having only one (\mathcal{T}_2 and $\mathcal{T}_2^{\text{ev}}$), thought the difference is much smaller. In addition, we observe that there is a very little difference between performance of the solutions from \mathcal{T}_1 and $\mathcal{T}_2^{\text{ev}}$, most likely due to the fact that the end-of-horizon issues of the former are mitigated by using it in a rolling-horizon fashion.

Table 7: Solution to the out-of-sample test in Section 3.3. The table shows the cost in the first two stages plus the total cost for solutions computed using four different tree structures, all evaluated on tree \mathcal{T}_4 .

Solution from	cost 1+2 (€)	obj. value (€)
deterministic	469 110	1 012 065.0
tree \mathcal{T}_1	458 602	907 827.0
tree \mathcal{T}_2^{ev}	460 951	905 505.0
tree \mathcal{T}_4	494 894	900 883.5

3.4 Effect of extra periods on a model with two proper stages

This test is analogous to the one presented in Section 3.1, except that we consider models with two proper stages. Again, we want to determine how many *extra-periods* we have to add to the scenario tree \mathcal{T}_4 to mitigate the end-of-horizon effects. For this, we introduce two extra scenario-tree structures \mathcal{T}_5 (see Figure 6) and \mathcal{T}_6 , obtained by adding respectively one and two extra-periods to \mathcal{T}_4 .

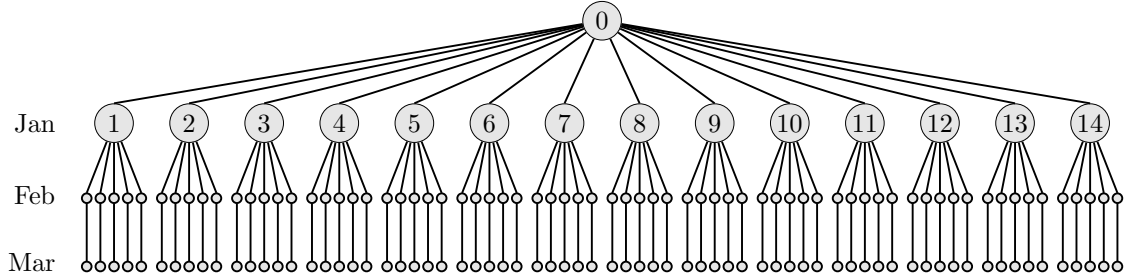


Figure 6: Scenario tree \mathcal{T}_5 .

The results of the tests are reported in Table 8. Just as in Section 3.1, we can see that the first-stage solution X_i^0 stabilizes just by adding *one extra-period* to the scenario tree \mathcal{T}_4 . And just as before, the difference between the *stable* solution and the solution from tree \mathcal{T}_4 is that the former has slightly more expensive first-stage solution—even though this time the number of vehicles does not change, since it was on its upper bound already in the case of tree \mathcal{T}_4 . (However, additional tests with higher value of the initial inventory level l_0 show that tree \mathcal{T}_5 indeed leads to more booked vehicles in the cases where tree \mathcal{T}_4 produces orders below the upper bound.) We may also notice that even though the trees \mathcal{T}_5 and \mathcal{T}_6 lead to equal first-stage solution, the latter has a more expensive second-stage solution. This is caused by a very minor end-of-horizon effect in the model with tree \mathcal{T}_5 , an effect analogous to what we observed in Section 3.1.

Table 8: Solution to the test in Section 3.4. For each scenario tree considered, the table shows the number of scenarios S and stages T , the optimal number of booked vehicles X_i for each supplier $i \in \mathcal{I}$, and the optimal costs in the root, the first two and first three stages, respectively.

tree	S	T	AG	CS	PA	VV	cost 1 (€)	costs 1+2 (€)	cost 1+2+3 (€)
\mathcal{T}_4	70	3	426	21	603	210	538 101	494 894	900 883
\mathcal{T}_5	70	4	417	30	603	210	543 933	498 782	934 058
\mathcal{T}_6	70	5	417	30	603	210	543 933	498 853	936 115

3.5 Sensitivity with respect to parameter α

In this section, we investigate the sensitivity of the optimal solution to the cancellation fee $\alpha \in [0, 1]$, for two different scenario tree structures. We start with results of the sensitivity analysis on the two-stage model with scenario tree \mathcal{T}_1 , presented in Table 9.

Table 9: Sensitivity analysis with respect to α , tested on scenario tree \mathcal{T}_1 . For each value of α , the table presents the optimal number of booked vehicles for each supplier $i \in \mathcal{I}$, the total number of booked vehicles and the optimal objective value, i.e. the minimal cost.

α	AG	CS	PA	VV	total	obj. val.
0.0	401	60	563	236	1260	334 541
0.2	416	33	563	248	1260	384 550
0.4	400	0	563	117	1080	422 096
0.6	303	0	563	116	979	451 662
0.8	303	0	537	110	950	476 524
1.0	300	0	533	110	943	500 484

We see that the optimal number of ordered vehicles decreases with α , while the total cost increases. While this might look counter-intuitive, it is a natural consequence of the way we model the booking and cancellation process: when the cancellation fee α is low, we can book vehicles that we are going to need only in some of the scenarios, as long as the expected price of buying extra clinker is higher than the expected cancellation fees. Alternatively, we can rewrite the objective function (1) as

$$\min \alpha q \sum_{i=1}^I t_i X_i + \sum_{s=1}^S p^s \left[b Y^s + (1 - \alpha) q \sum_{i=1}^I t_i Z_i^s \right], \quad (1')$$

which suggests a different (but equivalent) interpretation of the model: we pay the cancellation fee upfront and then pay the rest of the price for the vehicles we actually use. From this formulation it is easy to see why small values of α should lead to bigger values of X_i 's.

Let us for a moment consider the extreme case of $\alpha = 0$, i.e. zero cancellation fees. From (1'), we immediately see that the objective function in this case becomes

$$\min \sum_{s=1}^S p^s \left[b Y^s + q \sum_{i=1}^I t_i Z_i^s \right],$$

that is the first stage completely disappears from the model. In other words, there are no longer any decisions done under uncertainty (in the first stage). It follows that the case $\alpha = 0$ is equivalent to the so-called *wait-and-see* (WS) solution corresponding to models with $\alpha \in [0, 1]$, see Birge and Louveaux (1997) or Kall and Wallace (1994) for more information. This in turn allows us to compute the *expected value of perfect information* (EVPI), given by the difference between the objective values of the stochastic and wait-and-see solutions. In the case of $\alpha = 0.5$, used in all the previous tests, we get

$$\begin{aligned} \text{EVPI} &= \text{obj. val.}(\mathcal{T}_1) - \text{obj. val.}(WS) \\ &= 438\,404 - 334\,541 = 103\,763. \end{aligned}$$

This means that we should be willing to pay as much as a hundred thousand euro for the possibility to postpone the orders til after we have observed the actual demand and production level. Note that this

is almost a quarter of the total costs, suggesting that the company could save significant amounts if it could postpone the orders, or at least improve its estimates of the stochastic parameters.

To check whether the above results hold also in the multiperiod case, we have repeated the tests for scenario tree \mathcal{T}_5 . As we can see from Table 10, the pattern is the same as above, that is the objective value increases with α , while the number of vehicles booked in the first stage decreases. However, the case of zero cancellation costs no longer corresponds to the *wait-and-see* solution, as zero cancellation costs allow postponing the decision one period, while the WS model implies postponing all the decisions to the last stage of the model.

Table 10: Sensitivity analysis with respect to α , tested on scenario tree \mathcal{T}_5 . For each value of α , the table presents the optimal number of booked vehicles for each supplier $i \in \mathcal{I}$, the total number of booked vehicles, the cost in the first two stages and the optimal objective value, i.e. the total cost.

α	AG	CS	PA	VV	total	cost 1+2	obj. val.
0.0	425	33	592	210	1260	389 186	1 102 282
0.2	433	0	603	224	1260	435 228	1 174 898
0.4	426	21	603	210	1260	476 651	1 240 520
0.6	372	30	630	210	1242	510 520	1 303 512
0.8	316	33	563	163	1075	513 426	1 356 732
1.0	303	33	563	133	1032	525 885	1 397 606

4 Conclusions

We have proposed a two-stage and a multistage stochastic models for the stochastic single-sink transportation problem and tested it on data provided by the biggest Italian cement producer. Our tests show that for trees with one or two *proper stages*, it is enough to add one extra-period to mitigate the undesirable end-of-horizon effects. Furthermore, we have showed the advantage of using a stochastic model, compared to a deterministic one, as well as the additional advantage of having a model with more than one proper stage. Finally, a sensitivity analysis of the optimal solutions to value of cancellation fee revealed the importance of this parameter; In addition, the tests allowed us to compute the *expected value of perfect information* (EVPI) and thus show that a good estimate of the stochastic parameters could lead to significant savings for the company.

Acknowledgements

The authors would like to thank the whole *Italcementi* Logistic Group, in particular Dott. F. Gervasoni, P. Margorani and S. Nannini who helped us to understand the problem and provided us with the historical data. Furthermore, we are grateful to Elisabetta Allevi, Marida Bertocchi and Stein W. Wallace for helpful suggestions on a preliminary draft of the paper.

References

- B. Alidaee and G.A. Kochenberger. A note on a simple dynamic programming approach to the single-sink, fixed-charge transportation problem. *Transp. Sci.*, 39(1):140–143, 2005. ISSN 0041-1655 (print), 1526-5447 (online). doi: 10.1287/trsc.1030.0055.

- J. R. Birge and F. Louveaux. *Introduction to stochastic programming*. Springer-Verlag, New York, 1997. ISBN 0-387-98217-5.
- J. Dupačová. Stochastic programming: Approximation via scenarios. In *Proceedings of the 3rd International Conference on Approximation and Optimization in the Caribbean*. Benemérita Universidad Autónoma de Puebla, Puebla, Mexico, 1997. URL <http://www.emis.de/proceedings/3ICAOC/index.html>.
- J. Dupačová, G. Consigli, and S.W. Wallace. Scenarios for multistage stochastic programs. *Ann. Oper. Res.*, 100(1–4):25–53, 2000. ISSN 0254-5330 (print), 1572-9338 (online). doi: 10.1023/A:1019206915174. Research in stochastic programming (Vancouver, BC, 1998).
- Y. T. Herer, M. J. Rosenblatt, and I. Hefter. Fast algorithms for single-sink fixed charge transportation problems with applications to manufacturing and transportation. *Transp. Sci.*, 30:276–290, 1996. ISSN 0041-1655 (print), 1526-5447 (online).
- P. Kall and S.W. Wallace. *Stochastic Programming*. Wiley, Chichester, 1994.
- M. Kaut and S. W. Wallace. Evaluation of scenario generation methods for stochastic programming. *Pac. J. Optim.*, 3(2):257–271, 2007. ISSN 1348-9151.
- B. W. Lamar and C. A. Wallace. Revised-modified penalties for fixed charge transportation problems. *Manage. Sci.*, 43(10):1431–1436, 1997.
- B.W. Lamar, Y. Sheffi, and W.B. Powell. A capacity improvement lower bound for fixed charge network design problems. *Oper. Res.*, 38(4):704–710, 1990.
- U.S. Palekar, M.H. Karwan, and S. Zionts. A branch-and-bound method for the fixed charge transportation problem. *Manage. Sci.*, 36(9):1092–1105, 1990.